10-2018

Unearthing the X-Streams: Visualizing Water Contamination

Akangsha BANDALKUL
Singapore Management University, akangshab.2016@mitb.smu.edu.sg

Angad SRIVASTAVA
Singapore Management University, angads.2016@mitb.smu.edu.sg

Kishan Bharadwaj SHRIDHAR
Singapore Management University, kishanbs.2016@mitb.smu.edu.sg

Jason Guan Jie ONG
Singapore Management University, jason.ong.2016@mitb.smu.edu.sg

Yanrong ZHANG
Singapore Management University, yrzhang.2016@mitb.smu.edu.sg

Unearthing the X-Streams: Visualizing Water Contamination

VAST Mini-Challenge 2 Honorable Mention: Clarity of Narrative

Akangsha Bandalkul, Angad Srivastava, Kishan Bharadwaj Shridhar, Ong Guan Jie Jason, Zhang Yanrong*
Singapore Management University

ABSTRACT

The datasets released for VAST 2018 Mini Challenge 2 pertain to sensor readings capturing chemical concentrations and physical properties from water bodies in the Boonsong Lekagul wildlife preserve. This challenge is in continuation to the VAST 2017 Challenge, where the company Kasios was identified as the culprit in dumping the chemical - Methylosmoline. In the absence of actual chemical measurements in the soil, challenge participants need to visualize chemical contamination based on the proximal water bodies to identify trends of interest. A horizon plot developed helps to narrow down the complete list of 106 chemicals provided to only 7, from where control charts are used to identify change points and periods of abnormal activity. In addition, a sequential sampling strategy to capture differences in chemical patterns along the flow of the river is recommended using a sunburst diagram.

Keywords: Hydrology, Spatiotemporal analysis, Horizon plot, Sunburst diagram.

1 INTRODUCTION

The paper talks about our findings towards the VAST Challenge 2018, Mini Challenge 2. It is organized by the VAST committee, as part of the IEEE VIS conference. We explain the choice of our tools, the specific visualizations developed and significant findings that we discovered as part of our submission. The released dataset contains the readings of 106 chemicals, which were measured at 10 different locations across a wildlife preserve. The temporal range of the data spans over 19 years of data. A comprehensive approach to identify the set of chemicals which might be of interest was first performed. Following this, a visualization of the chemical trends over time was done. The team also identified a few avenues where the sampling strategy of the hydrology department which provided these datasets could be improved. The Tableau workbook showing our visualisations can be accessed from this link.

2 THEORY

The section talks about the impetus behind using the tools of our choice, Tableau and d3.js.

2.1 Tableau

With 19 years of data present, we needed ways where we could identify macro level trends, and then narrow down to specific time ranges such as yearly and monthly trends. With features such as filters and pages, Tableau helped us accomplish this.

* {akangshab.2016, angads.2016, kishanbs.2016, jason.ong.2016, yrzhang.2016}@mitb.smu.edu.sg

2.2 d3.js

d3.js was used for the sunburst diagram with which the team developed and identified a sampling strategy which the hydrology department could use. The flexibility of being allowed multiple tools to work on the challenge meant we could tailor each of our visualizations using the most appropriate tool.

3 VISUALIZATION TYPES

The section describes the ideology behind the usage of specific visualization types in answering the questions of the mini challenge.

3.1 Horizon Plot

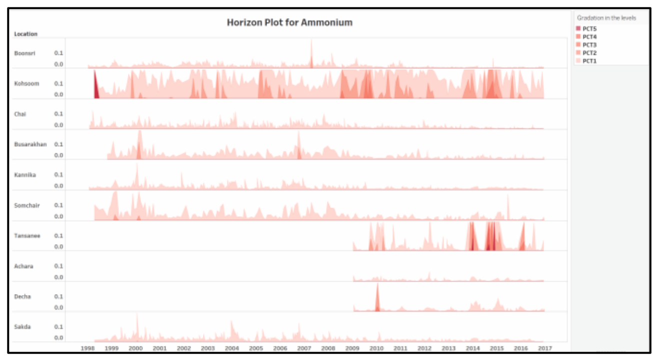

Figure 1: The Horizon Plot for Ammonium

We use the horizon plot [1] to provide a unified view for a chemical along both spatial (10 locations) and temporal (228 months) dimensions. In the figure above, the peak reading of Ammonium occurs at Kohsoom in 1998, with Tansanee showing spikes of interest in recent years.

Discussed below are a few considerations that went in developing the visual.

3.1.1 Min-Max scaling

Min-max scaling was performed to handle the natural variation in the range of the readings of chemical concentrations. For e.g. the maximum Ammonium reading was 14.87 mg/l whereas for Total Dissolved Phosphorus, it was 0.57 mg/l. This variation is naturally expected. There was also a difference in their units of measurement.

3.1.2 Adjustable bin-width

After scaling from 0 to 1, for some chemicals, it was found there were lot of measurements which were 0. This could have been natural or due to faulty calibration of measuring equipment. However, since no concrete information was available, an adjustable bin-width parameter was introduced, to show the different gradation levels PCT1 to PCT5.

3.2 Control Chart

From the filtered list of chemicals identified with the horizon plot, control charts were used on the actual values of chemical readings to examine in greater detail along the temporal dimension.

3.2.1 Reason for animation

We leveraged Tableau's pages in capturing snapshots of a chemical level at a point in time. Animating the set of snapshots, we were able to produce visualizations for the set of chemicals we identified with the horizon plot showing exact periods of abnormal activity. This explains the reason for the usage of multiple GIF files in the submission.

3.3 Cycle Plot

A cycle plot is useful for trends that are seasonal in nature. Water temperature and Dissolved Oxygen, which are inversely proportional to each other, exhibit a cyclical pattern. The plots showed us the months where the temperatures were significantly higher were June, July and August.

3.4 Sunburst Chart

The sunburst chart is a useful tool which helps to visualize sequential data. We converted the dataset into a format from where a sequence could be studied.

3.4.1 Identifying the sequence in the dataset

The below table shows how the dataset can be made into a sequential form, suitable for showing scattered hierarchy in the form of a sunburst diagram. The complete sequence for all 228 months (19 years x 12 months), using a python code snippet. A quick sanitary check yielded the finding that the hydrology department did not return to the same location twice in a given month. This paved the way for us to develop a visualization to see the sequence in which the department went about sampling the various locations in the preserve. Figure 2 shows one such sequence for a river system, indicating that the measurements were not coherent.

Table 1: Converting measurement data into a sequence

<table>
<tr><td>Time</td><td>Location</td><td>Resultant sequence for Jan 1998</td></tr>
<tr><td>11th Jan 1998</td><td>Boonsri</td><td rowspan="4">Boonsri
Kannika
Chai
Kohsoom</td></tr>
<tr><td>18th Jan 1998</td><td>Kannika</td></tr>
<tr><td>29th Jan 1998</td><td>Chai</td></tr>
<tr><td>31st Jan 1998</td><td>Kohsoom</td></tr>
</table>

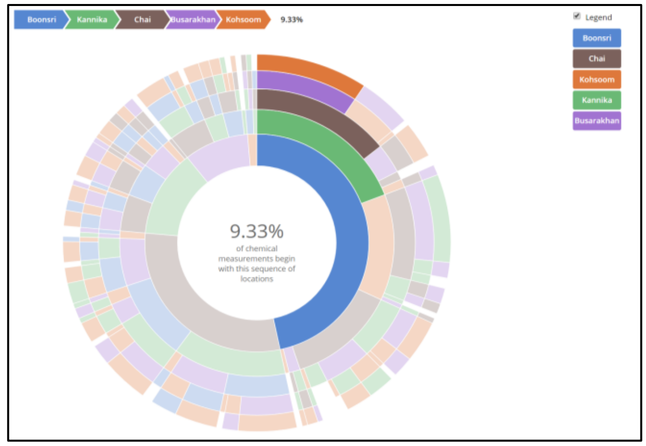

Figure 2: The sunburst diagram shows that this particular sequence is followed by the department 9.3% of the time (21 months between Jan 1998 and Dec 2016)

4 Recommendation Of A Sampling Strategy

The first step in recommending a sampling strategy was to arrive at an understanding of their existing sampling strategy, and the potential outstanding pitfalls in them.

4.1 Identifying the sampling frequency

We define the sampling frequency as the inverse of the time between two consecutive measurements at a particular location. This location is independent of the chemical measured, i.e. if 100 chemicals were measured in Boonsri in Jan 1998, and 50 different set of chemicals were measured in Oct 1998, we define the time between the measurements to be 7 months. The days between consecutive samples showed clear patterns at certain locations, from where we discovered that the sampling frequency was incoherent.

4.2 A possible route for sampling

We believed measurements taken upstream of the river could be aligned at a fixed frequency to the downstream of the river. This would help to identify chokepoints in the river and monitor if there were any critical regions of contamination. The preserve map helped us to illustrate this clearly in the video.

5 Conclusion

We leverage Tableau and d3.js to show how various visualizations can be developed for handing spatiotemporal data. In doing so, we also uncovered how the hydrology department has performed its sampling over the years. Our key recommendation was that to effectively analyze contamination, monitoring upstream changes at regular intervals in downstream was pivotal. To facilitate this, the chemicals of interest may be narrowed down using techniques such as the horizon plot and the control chart.

References

[1] Y. Fornes, (2016, April 5). How to create a horizon chart to display dense data Retrieved June, 2018, from https://www.tableau.com/about/blog/2016/4/visualizing-dense-data-how-cut-and-superpose-areas-52839